Robotically aligned optical coherence tomography with 5 degree of freedom eye tracking for subject motion and gaze compensation

PABLO ORTIZ,[1,*] **MARK DRAELOS,**[1] **CHRISTIAN VIEHLAND,**[1] **RUOBING QIAN,**[1] **RYAN P. MCNABB,**[2] **ANTHONY N. KUO,**[2,1] **AND JOSEPH A. IZATT**[1,2]

[1]*Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA*
[2]*Department of Ophthalmology, Duke University, Durham, NC 27708, USA*
[*]*pablo.ortiz@duke.edu*

Abstract: Optical coherence tomography (OCT) has revolutionized diagnostics in ophthalmology. However, OCT requires a trained operator and patient cooperation to carefully align a scanner with the subject's eye and orient it in such a way that it images a desired region of interest at the retina. With the goal of automating this process of orienting and aligning the scanner, we developed a robot-mounted OCT scanner that automatically aligned with the pupil while matching its optical axis with the target region of interest at the retina. The system used two 3D cameras for face tracking and three high-resolution 2D cameras for pupil and gaze tracking. The tracking software identified 5 degrees of freedom for robot alignment and ray aiming through the ocular pupil: 3 degrees of translation *(x, y, z)* and 2 degrees of orientation *(yaw, pitch)*. We evaluated the accuracy, precision, and range of our tracking system and demonstrated imaging performance on free-standing human subjects. Our results demonstrate that the system stabilized images and that the addition of gaze tracking and aiming allowed for region-of-interest specific alignment at any gaze orientation within a 28° range.

1. Introduction

Optical coherence tomography (OCT) [1] is a non-invasive imaging modality that has revolutionized structural imaging of *in vivo* biological tissues in ophthalmology by providing images with micrometer scale resolution of the anterior [2,3] and posterior [4,5] segments of the eye. As a result, OCT has become a standard technology employed to manage ocular diseases [6,7]. Most OCT systems, however, are large tabletop devices exclusively found in imaging rooms at eye care offices. Additionally, these systems require skilled ophthalmic technicians to operate and align as well as cooperative patients who must be capable of sitting upright and follow directions to use a chinrest and direct their gaze at a fixation target. Consequently, OCT is not readily accessible for urgent and routine care environments, such as primary care clinics.

The most common approaches thus far to solve these accessibility problems are through a combination of handheld OCT probes [8–15] and image registration techniques [15–22]. Handheld probes provide the workspace flexibility to image subjects outside an office setting by allowing an operator to bring the OCT scanner to non-seated subjects. Additionally, retinal OCT probes provide the maneuverability that permits imaging from various angles of entrance, allowing the operator to image the desired region of interest without the need of a fixation target. Handheld probes have great potential to increase the accessibility of OCT; however, their image quality is limited by the stability of the operator handling the probe. Digital motion correction, in the form of image registration, helps correct motion artifacts from operator tremor and/or patient

motion. However, these image registration methods cannot correct large-scale motion artifacts that lead to a loss of data due to misalignment, thus there is still a need for highly trained and skilled technicians with steady hands. Handheld devices are a good solution for the workspace and patient cooperation barriers of OCT accessibility, but are still limited by the operator skill barrier.

In an effort to ease the operation of OCT systems, commercial manufacturers have incorporated limited eye-tracking and hardware-based self-aligning mechanisms into tabletop scanners [23–25]. These systems are much easier to operate than handheld probes, but still require chin and forehead rests, lack complete workspace flexibility, and are still tabletop-bound, meaning they cannot be used to image non-seated patients such as those who are bedbound or free-standing. They also lack the maneuverability of handheld probes, meaning they rely on cooperative patients and fixation targets to image a region of interest. Ideally, one could integrate the hardware components that facilitate use of these systems into handheld probes to overcome both operator skill and workspace barriers. However, the addition of automated alignment components would increase the weight and size of the handheld probe, making it even harder to maneuver.

To overcome the aforementioned barriers, we previously developed and demonstrated a completely automated 3 degrees of freedom robotically-aligned system that aligns an OCT scanner to the subject and, through active motion tracking of their pupil, corrects for motion without any operator involvement [26,27]. By utilizing a robotic arm, we achieved high quality imaging in a large workspace. The increased workspace enabled imaging of non-seated subjects, as they did not need to position themselves within a small workspace, as they must do with tabletop OCT scanners. Additionally, due to the increased weight/fatigue tolerance of a robotic arm compared to a human operator, we were able to integrate self-alignment hardware components without compromising maneuverability. The system corrected for translational motion in 3D with accuracy and bandwidth large enough to stabilize the OCT images. Here we further enhance the system, increasing the degrees of freedom from 3 to 5 by integrating gaze tracking and rotational capabilities for precise control of scan angle of entrance into the eye. This advance allows imaging any retinal region of interest within the gaze tracking range without a fixation target. The new design also provides the ability for independent control of the imaging region on the retina and pupil entry position.

Gaze tracking—the problem of estimating the gaze orientation of the eye—has traditionally been addressed with the goal of studying the attention of subjects during human computer interactions in a variety of platforms. Multiple methods have been implemented in an attempt to solve the gaze tracking problem under different contexts, and they primarily consist of video-based systems that track the pupil of the subject. Two-dimensional regression methods that identify pupil contours and corneal reflections (PCCR) have been used to estimate gaze using simplistic single camera setups [28–31]; however, these methods assume little to no translational motion of the eye. Cross-ratio methods use projective transformation matrices to map the camera image space to the illumination source space [32,33], but they are vulnerable to axial motion from the subject [34]. Appearance based methods, which use the shape and texture properties of the eye, can also estimate gaze with single camera setups, even without additional illumination sources [35–37]. Shape based methods, where parabolic and circular templates are cross correlated with eye contour and iris, respectively, can estimate gaze with low-resolution cameras and no additional illumination sources [38,39]. Both appearance and shape based methods are highly portable since they don't require illumination LEDs or more than one camera, but their accuracy is lower than methods with calibrated illumination. 3D model-based methods use a geometric model of the eye to model the physical behavior of rays propagating from calibrated LEDs [40–42]. The most robust, precise, and accurate methods—which even tolerate head motion—are 3D based methods in calibrated multi camera and LED setups [34]. In this work, we leverage the high accuracy, precision, and robustness of 3D based gaze methods to enable motion and

gaze tracking with our robotically aligned OCT scanner, and demonstrate the performance of the system both in controlled model settings and in human subjects.

In this manuscript, we present a robotically aligned OCT system with 5 degree of freedom tracking for motion and gaze compensation. This is, to the best of our knowledge, the first OCT system that compensates translational and rotational motion without the need of a chinrest or fixation target. We aim to utilize this system to lower accessibility barriers to retinal OCT imaging by allowing motion-corrected imaging of multiple retinal regions of interest without the mechanical stabilization requirements of a chinrest, fixation target, or seating chair. It is important to lower these mechanical stabilization requirements in order to increase OCT access to patients who cannot physically utilize a standard OCT patient interface, as well as to minimize the required skillsets necessary to operate an OCT scanner.

2. Methods

2.1. OCT engine

The system's scanner used a $100kHz$, $1060nm$ swept source laser (Axsun Technologies Inc., Bilerica, MA) illuminating a Mach-Zehnder interferometer with a $2.7mm$ $1/e^2$ waist diameter at the subject's pupil plane. The interference signal was detected with $1.0GHz$ balanced receiver (Thorlabs, Newton, NJ) and digitized at $800MSamp/sec$ (ATS-9360, Alazar Technologies Inc., Pointe-Claire, Quebec). We optimized the optical design with OpticStudio (Zemax, Kirkland, WA) with a theoretical full width at half maximum (FWHM) point spread function (PSF) of less than $9\mu m$ (Figs. 1(C) and 1(D)) and ideal axial resolution of $5.4\mu m$ within a 20° field of view and imaging depth of $8.77mm$ at the retina. We imaged volumes at a rate of 1.75 seconds per volume, 200 B-Scans per volume, and 500 A-Scans per B-Scan.

2.2. Robot-mounted scanner design

The sample arm of our scanner utilized a 4F relay telescope to image a pair of galvanometer scanners to the subject's pupil (Fig. 1(B)). The telescope was designed with $2in$ diameter lenses to provide a comfortable working distance from the tip of the scanner to the patient cornea of $100mm$. We folded the optical path at the retinal conjugate plane of the 4F system using a fast steering mirror (FSM) (Optics In Motion LLC., Long Beach, CA). Additionally, we utilized folding mirrors between the FSM and both the relay and objective lenses to reduce the form factor of the scanner.

The scanner sample arm was also designed to support an eye tracking system that consisted of three pupil cameras (BFS-U3-042M, FLIR Systems): one camera was inline with the pupil (with an optical axis coaxial to that of the objective) while the other two were offset, each of them to one side of the objective (Fig. 1(B)). The offset pupil cameras were lowered $33mm$ relative to the original design in [26], such that they did not obstruct the line of sight of the non-imaged eye. In order to image with the inline pupil camera, we imaged through the OCT scanner's objective lens. To integrate the inline pupil camera's optical path with that of the OCT scanner, the fold mirror immediately behind the objective was a short-pass dichroic mirror with a $930nm$ cutoff that reflected the OCT beam and allowed other light to enter the pupil camera. To reduce the effect of room light flicker, the pupil cameras utilized a $850nm$ band-pass filter rejecting visible light. We illuminated the subject's pupil with four $850nm$ LEDs (Fig. 1(A)) to have enough signal in the pupil cameras. These LEDs were powered with a DC power supply (Korad Technology, Shengzhen, China).

To mount our optics to the robot, we designed 3D printed opto-mechanical housings. The parts were designed with CAD modelling software (Autodesk Inventor) and 3D printed with a Form 2 printer (FormLabs, Somerville, MA). The scanner, weighing $2.65kg$, was mounted on the robotic arm (Fig. 1(A)), which had a $3kg$ payload limit.
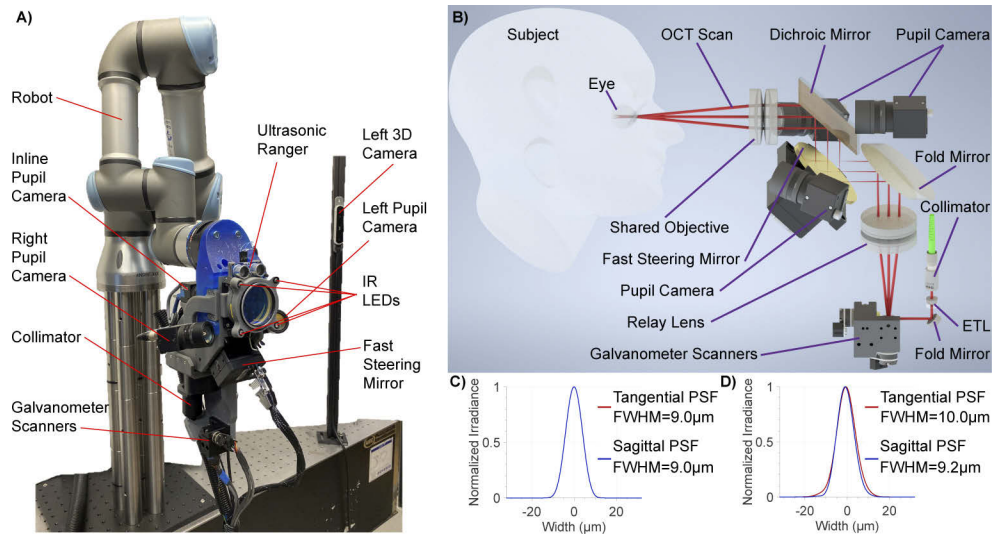
Fig. 1. Robotically Aligned Scanner Design. A) OCT scanner mounted on a UR3 robot. The left and right (not shown) 3D cameras were used to track facial landmarks of the subject. The pupil cameras were used to track the pupil with higher resolution and larger bandwidth than those of the 3D face cameras. The IR LEDs illuminated the field of view for the pupil cameras and generated Purkinje reflections used in gaze tracking. The Ultrasonic Ranger was not utilized in the methods of this manuscript. B) OCT scanner optical diagram. The scanner utilized a 4F telescope to image galvanometer scanners into the subject's pupil. The Fast Steering Mirror was optically conjugate to the retina. The Dichroic Mirror integrated the optical path of both the OCT Scan as well as the Inline Pupil Camera. The OCT beams at the extremes and center of a scan are shown in red. C) and D) show the tangential and sagittal point spread function at the retinal plane measured at the center $(0° \times 0°)$ and near the edge of the field of view $(8.5° \times 8.5°)$, respectively. We measure and report the FWHM value for the theoretical lateral spot. ETL: Electrically Tunable Lens.

2.3. Robotic alignment

We used a UR3 robotic arm (Universal Robots, Odense, Denmark) (Fig. 1(A)) to position the retinal OCT scanner such that the objective was located a working distance of $100mm$ away from the pupil of the subject and was coaxial with the region of interest. We controlled the robot at a $125Hz$ frequency, sending a new motion command every cycle. The workspace was restricted to a $300mm$ wide, $150mm$ deep, and $550mm$ tall virtual box where the subject could keep their face and remain within reach of the robotic arm during Face Tracking mode (Fig. 2). In addition to the spatial limits, we restricted the robot's velocity and acceleration limits. Linear velocity and acceleration limits were set to $10\frac{cm}{s}$ and $5\frac{cm}{s^2}$, respectively. Acceleration and velocity limits were enforced using an on-line time-optimal trajectory planning algorithm [43]. Angular velocity was limited to $\frac{\pi}{20}\frac{rad}{s}$ and was implemented using quaternion spherical linear interpolation during angular path planning. The robot was controlled through a state machine shown in Fig. 2. During Face Tracking mode, an open-loop control system was implemented where the face tracking cameras identified the 3D eye location in the robot base coordinate system and the robot moved the scanner to a fixed offset from this point. During Pupil Tracking mode, a closed-loop system was implemented where the location of the pupil relative to the pupil cameras was measured. Since the pupil cameras were mounted onto the robot arm (Fig. 1(A)), they were not fixed in the robot base coordinate system as the arm moved during a session. The robot then moved to correct for any error between the measured pupil location and the objective's back focal point,

whose location was pre-calibrated and was fixed relative to the pupil camera coordinate system. For any given region of interest on the retina, we define the axis of interest: the axis that the objective lens needs to be coaxial to in order to image the region of interest. The axis of interest for a given region is at a fixed angular offset from the eye's optic axis. During Gaze Tracking mode, the optic axis of the eye was measured relative to the pupil camera coordinates. The axis of interest was then calculated by adding its pre-calibrated angular offset to the measured optic axis. From the axis of interest, two angles were extracted: *yaw* and *pitch* relative to the pupil camera coordinates. The target pose of the robot was subsequently rotated around the pupil by the *yaw* and *pitch* angles.
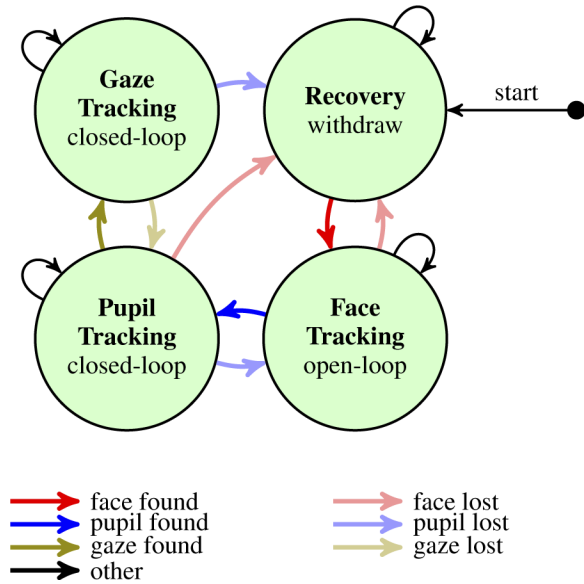

Fig. 2. State machine for robotic alignment during imaging session. The robot began the session in *Recovery* mode, where it remained still in its resting position until the 3D cameras identified a face. If *Recovery* mode was entered at a later stage of the session, the robot withdrew from its current location towards resting position. During *Face Tracking* mode, the robot utilized eye location information from the 3D cameras and moved such that the objective of the scanner was 100*mm* in front of the eye. The state machine transitioned to *Pupil Tracking* once it found a pupil. During *Pupil Tracking* mode, the robot utilized information from the pupil cameras to finely correct its location. If gaze orientation was successfully identified, the state machine entered *Gaze Tracking* mode, where the robot pivoted the scanner around the subject's pupil such that the optical axis of the scanner was aligned with the axis of interest.

2.4. Eye tracking

We utilized 30*Hz* tracking of the subject's face to provide an initial estimation of eye location [26]. We used two 3D cameras (RealSense D415, Intel), each mounted to the table and on each side of the robotic arm (Fig. 1(A)), imaging the entire workspace of the scanner. The pose for each 3D camera was calibrated utilizing a hand-eye calibration procedure [45] based on images acquired from an ArUco target [46] mounted on the robot end-effector. We detected facial landmarks (Figs. 6 and 7) with the cameras' infrared images using OpenFace 2.0 [47] and used them to extract the 3D position of both eyes.

Additionally, we used the pupil cameras operating at 200*Hz* to track the eye imaged during the OCT session (Fig. 7). The left and right pupil camera poses were calibrated relative to the

inline pupil camera's coordinate frame by using chessboard calibration targets [48]. Each camera was calibrated using OpenCV stereo calibration [49] and refined with bundle adjustment. We detected the pupil from each camera view using custom C++ software that identified dark circular connected components within the image [27,44]. We estimated the pupil position in 3D using linear triangulation from the detected images.

We estimated and tracked subject gaze at 200Hz to identify not only the location of the eye but also the orientation to co-localize the ocular and scanner optical axes. Our gaze tracking method was adapted from a general-purpose PCCR gaze tracking algorithm [42]. After finding the pupil center in each pupil camera view, our custom software identified the corneal reflections from the LEDs by finding saturated circular regions in the proximity of the pupil image center (Fig. 3). For each corneal reflection found, we defined a plane using 3 points in space: the camera's principal point (which was calculated through the calibration procedure mentioned in the previous paragraph), the corneal reflection image point at the sensor plane of the camera, and its corresponding light source (LED) location (Fig. 4). We then calculated the center of corneal curvature as the intersection point for all identified planes. The axis defined by the center of corneal curvature and the center of the pupil was recorded as the optic axis of the eye. The location of the illumination LEDs was extracted from a CAD model of the scanner. In order to match each corneal reflection image with its corresponding LED, we sorted all the identified reflections as well as 3D LED locations based on their angle relative to the horizontal axis.
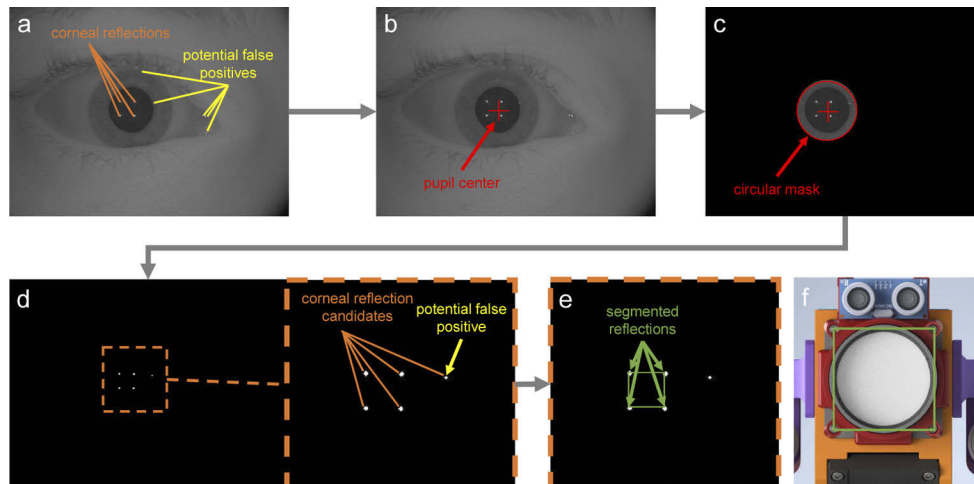

Fig. 3. Corneal Reflection Segmentation. a) typical pupil camera image. b) pupil camera image with red cross at the pupil center. The pupil center was identified as the center of the largest dark circle in the image [27,44]. c) To remove potential false positive reflections (yellow in panel a) located far from the pupil, we applied a circular mask centered at the pupil center. d) afterwards, a binary threshold excluded non-bright pixels and connected components was utilized to identify remaining corneal reflection candidates. e) in order to filter out remaining potential false positive reflections (yellow in panel d), we selected the set of candidates whose distribution best matched that of the IR LEDs in the scanner, which were distributed in a square pattern (shown in green in panel f).

2.5. Optical alignment

The robotic alignment procedure provided a large workspace over which to align the system. However, mechanical motion alone was insufficient to compensate for physiological motion [27]. To address this issue, our system used feedforward control from the pupil camera tracking
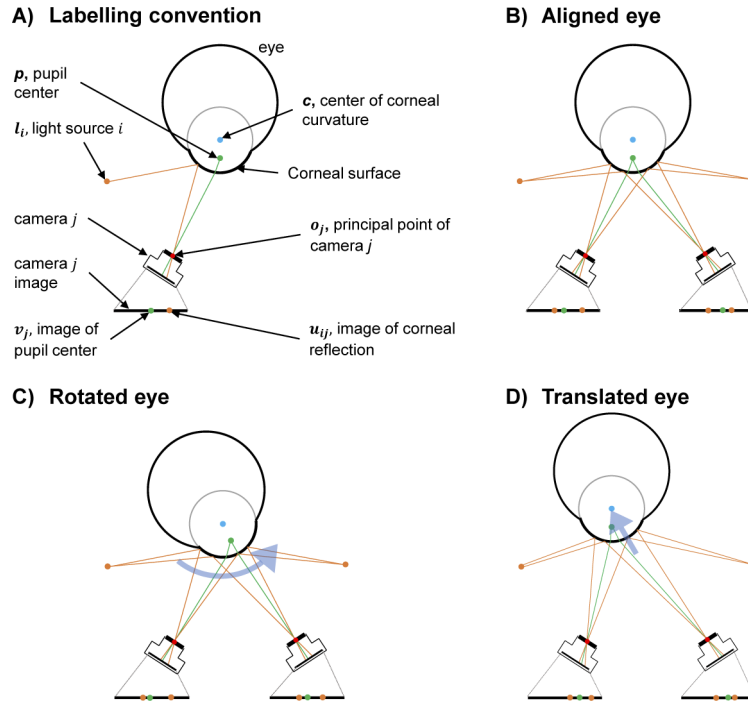
Fig. 4. Gaze Tracking algorithm. Pupil center images v_j for a rotated eye (C) are shifted relative to those from an aligned eye (B). Similarly, v_j and u_{ij} for a translated eye (D) are shifted from those in an aligned eye (B). The gaze tracking algorithm is summarized as: 1) Calibrate cameras and LEDs to find o_j and l_i. 2) Segment images to find v_j and u_{ij} as described in Fig. 3. 3) For each ij combination, define a plane cpl_{ij} containing l_i, o_j, and u_{ij}. 4) Find intersecting point between planes cpl through linear optimization to calculate c. 5) Calculate p via linear triangulation of v_j. 6) Calculate optical axis as the axis defined by points p and c.

to optically align the scan to the subject's pupil with sufficient bandwidth to maintain pupil centration (Fig. 5). To this end, a FSM was utilized to tilt the beam at the retinal conjugate plane, which shifted the beam's point of entrance at the pupil plane [44]. This allowed us to accurately correct for lateral motion from the subject over a large bandwidth. To optically correct for axial motion, we mounted a retro-reflector on a motorized stage to path length match the reference and sample arm as the subject moves axially. For optical correction of angular offset of the scan, we adjusted the OCT scan waveform using a summing amplifier inserted before the input of each galvanometer axis. The measured *yaw* and *pitch* angles from gaze tracking were converted to a voltage offset of the horizontal and vertical axis galvo mirror, respectively. Lateral, axial, and angular motion were therefore corrected simultaneously, and independently from each other at $6.2Hz$, $6.1Hz$, and $17.7Hz$ bandwidth, respectively [26].

2.6.   *Face tracking performance characterization*

In order to characterize the performance of our face tracking system, we mounted a Styrofoam head (Fig. 6) model onto a motorized translation stage and placed it inside the field of view of both face cameras. Once both cameras identified the Styrofoam head, we began tracking each of its eyes' location. We then stepped the stage in $3mm$ increments. We calculated precision by subtracting the average measurement at each step from the raw data and calculating the standard deviation of the resulting signal. Accuracy was evaluated by subtracting the ground truth, as
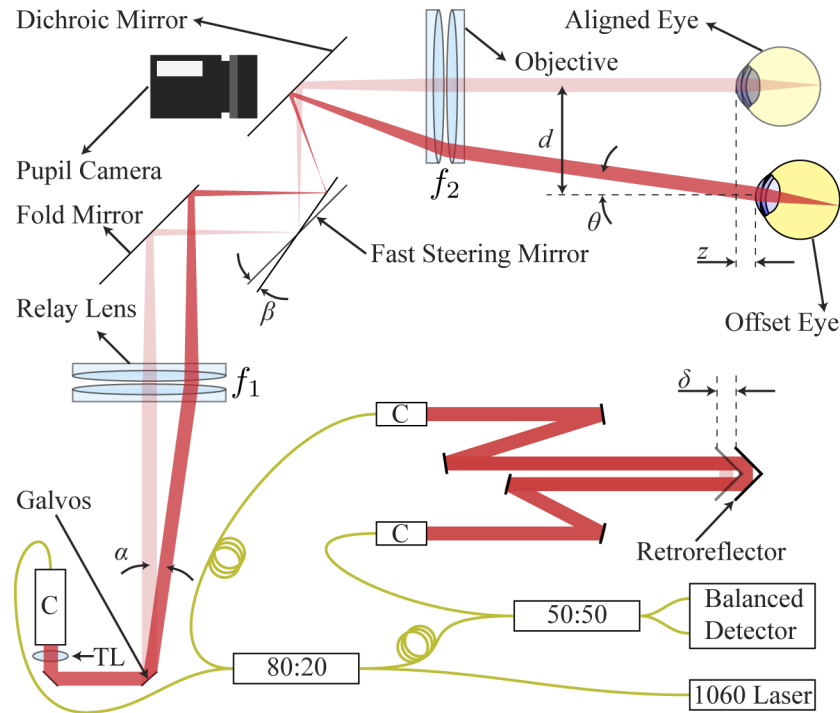
Fig. 5. Optical Alignment. The sample arm of the imaging system is shown imaging an aligned eye (transparent optical path) and an offset eye (non-transparent optical path). The offset eye is offset relative to the aligned eye by a lateral distance d, an axial distance z, and a rotational angle θ. The lateral displacement d is corrected by an angular tilt $\beta \approx \frac{d}{2f_2}$ at the fast steering mirror. The axial displacement z is corrected by an axial displacement $\delta = z$ at the retroreflector in the reference arm. The angular displacement θ is corrected by an angular offset $\alpha = \frac{f_2}{f_1}\theta$ added to the galvanometer waveform. C: transmissive collimator, TL: Tunable Lens.

measured from the translation stages, from the signal at each step and calculating the root mean squared (RMS) error of the resulting signal. We confirmed the face tracking range was at least as large as the virtual box that defined the workspace of the robot. To do this, we translated the Styrofoam head horizontally, vertically, and axially across the field of view of the cameras and confirmed the cameras tracked the face for at least the size of the corresponding dimension of the virtual box. The range of face tracking was then reported as the size of the virtual box. Even though the cameras have a field of view that goes beyond the virtual box, the range was reported only as the size of the virtual box because the robot was not allowed to move past this virtual box.

2.7. Pupil tracking performance characterization

To characterize translational tracking performance in axial and lateral motion, we mounted a pupil model on a motorized translational stage and moved it such that it was inside the field of view of all three pupil cameras and started tracking its location. We then stepped the stage in rapid 1*mm* increments until the pupil was no longer tracked by our system. This was done in both lateral and axial directions.

To characterize rotational tracking performance, we mounted the pupil model on a rotational stage and rotated it until the gaze tracking system was able to identify its gaze and started tracking it. We then stepped the stage in rapid 2° increments until the gaze was no longer tracked.
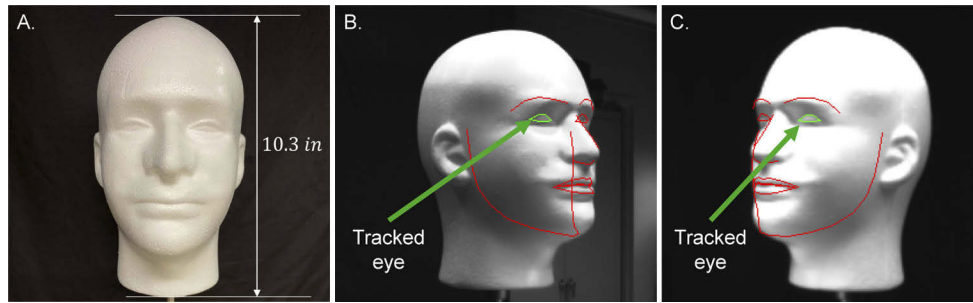
Fig. 6. (A) Styrofoam mannequin head utilized for face tracking performance characterization. Segmented images from the left (B) and right (C) face tracking cameras. Non-tracked facial landmarks are highlighted in red. Tracked facial landmarks for eye tracking are highlighted in green; these correspond to the right (B) and left (C) eye.

For each tracking subsystem, we evaluated precision, accuracy, and range. Precision was evaluated by subtracting the average measurement at each step from the raw data and calculating the standard deviation of the resulting signal. Accuracy was evaluated by subtracting the ground
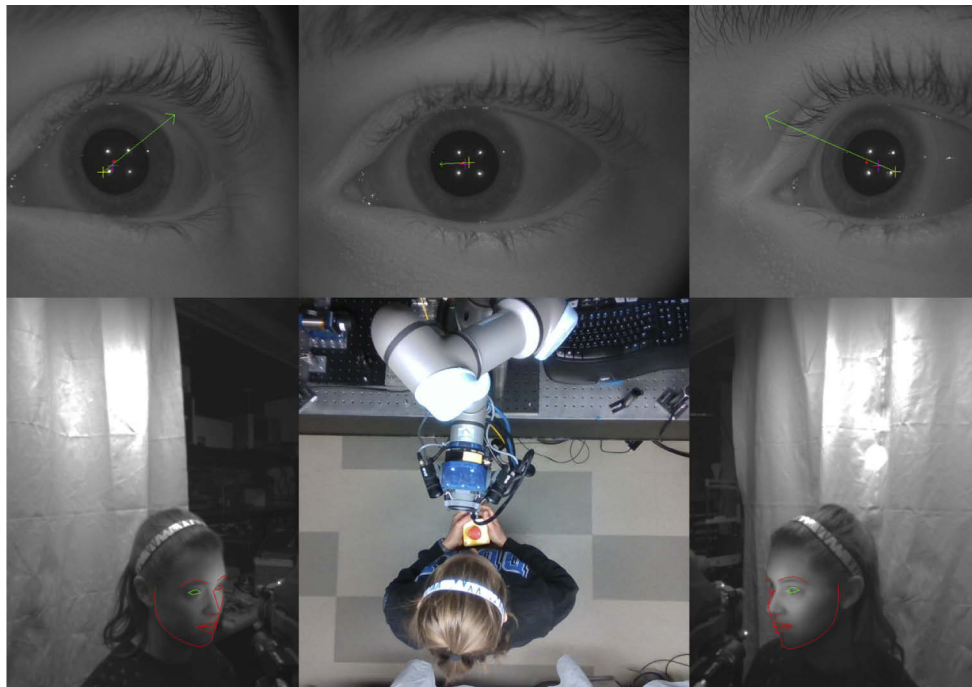

Fig. 7. Imaging session tracking system. The top three panels are images from the left, inline, and right pupil cameras. Shown is the projected optic axis (green arrow), target pupil location (magenta cross), measured center of corneal curvature (yellow cross), and segmented pupil center (red dot) during pupil and gaze tracking. The bottom panels are the Left 3D camera, an overhead camera (not used for tracking), and the Right 3D camera. The red outline in the left and right 3D cameras correspond to the segmented facial landmarks from the face tracking. The green segmentation corresponds to the corresponding eye that each 3D camera was tracking.

truth, as measured from the rotation and translation stages, from the signal at each step and calculating the root mean squared (RMS) error of the resulting signal. The range was evaluated as the range over which the reported accuracy and precision were measured.

2.8. Subject imaging

We imaged four healthy subjects, all between 20 and 30 years old. The subjects were consented under a Duke University Health System IRB-approved protocol in accordance with the Declaration of Helsinki. We followed the ANSI-Z136.1 laser safety standard, which led to a sample arm optical power of $1.59mW$ @$1060nm$ wavelength and an IR LED illumination power of $223\mu W$ at $850nm$ wavelength.

Subjects were asked to walk into the frame with an emergency stop button to stop the robot if desired. They were given no fixation target or mechanical head stabilization device. They were then asked to stand still while the robot aligned and imaged their eyes (Fig. 7).

The RAOCT system was then commanded to align to 4 regions of interest for each subject: right optic nerve, right fovea, left optic nerve, and left fovea. Volumes were collected at each region of interest and then the subject was dismissed.

Due to COVID19 safety protocols, subjects were required to wear masks at all times during the imaging session. In order to minimize the negative effect that facial masks had on the face tracking quality, subjects were given a secondary mask with realistic facial features printed on to them and asked to wear those on top of their personal masks. Additionally, an operator asked the subjects to look directly at each face tracking camera before the imaging session. This was done because the face tracking software utilized previous frame information in tracking a face during a video session, and it worked more robustly when the subject looked directly at the camera. This meant that reliable face tracking at the beginning of the imaging session improved face tracking during the rest of the imaging session, even if the subject was no longer looking directly at the cameras.

In addition to hardware motion correction, we registered the acquired volumes with B-Scan rigid-body transformations. The first B-Scan from the raster scan was chosen as a reference to register the rest of the volume to. Other B-Scans were registered by applying the rigid-body transformation [50] that maximized their cross-correlation with the adjacent, previously registered, B-Scan. This minimized any residual motion during acquisition. It is worth noting that without robotic alignment and hardware motion compensation, digital registration alone would be insufficient to correct for motion, as we would not be able to acquire the high enough quality raw data required for digital registration to work in the first place.

3. Results

3.1. Performance characterization

The tracking accuracy, precision, and range of our Face and Pupil Tracking are reported in Table 1. We achieved a gaze tracking accuracy of less than $\frac{1}{3}^{\circ}$ across the 28° range, over which 15 different angular orientations were evaluated, with an average of 848 data points recorded at each orientation.

The data from our Gaze Tracking performance experiments is shown in Fig. 8.

3.2. Subject imaging

In Fig. 9, volumes from each of the regions of interest imaged during the session are shown for 4 subjects imaged under the aforementioned protocol. All imaging sessions resulted in stabilized volumetric data of the target features of interest: optic nerve head and foveal pit for left and right eyes. All subjects were imaged successfully despite variations in gender, height, iris color, and ethnicity, and despite them wearing masks covering the bottom half of their faces.

Table 1. Eye Tracking Performance Characterization

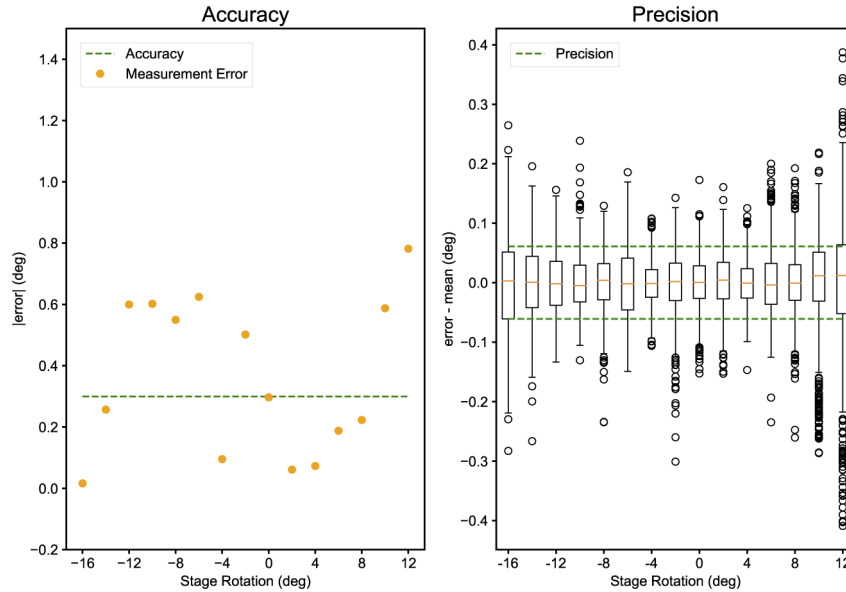|		Accuracy	Precision	Range
Face Tracking	Lateral Translation	4.1mm	0.3mm	300mm
	Axial Translation	9.7mm	0.11mm	150mm
Pupil Tracking	Lateral Translation	24.0μm	6.5μm	32mm
	Axial Translation	31.5μm	6.2μm	22mm
	Gaze Rotation	0.30°	0.06°	28°

Fig. 8. Gaze Tracking accuracy and precision. The panel on the left shows the absolute error of the mean measurement at each level of rotation (orange) as well as the root mean squared error (0.30°) over the reported field of view of the tracking system (green). The panel on the right shows the gaze measurement distribution at each level of rotation subtracting the mean measurement at each level. The reported precision (green) was 0.06°. The observed asymmetry in the angular range relative to 0° is due to the presence of the nose after rotating in one direction but not the other.

Figure 10 shows the results of Subject 2 presented as *enface* projections as well as B-scans in the same format as that used in ophthalmic photography: both eyes' optical nerve and macular images along with their circumpapillary and foveal B-scans, respectively. We selected Subject 2's data because it was the brightest among the subjects (as seen in Fig. 9), but the amount of remaining motion stability seen in Fig. 10 was typical of all subjects. This demonstrates the ability to automatically generate the most common scans performed in an ophthalmic photography session.

Visualization 1 shows a fully autonomous (without an operator) imaging session of a maskless subject imaged prior to the setting of COVID19 safety protocols. The system acquired volumetric data from all 4 regions of interest in less than 1 minute. Visualization 2 shows a semi-autonomous (with an operator) imaging session of a masked subject imaged. Despite the presence of a mask occluding prominent facial landmarks, face tracking was robust enough to track the subject and successfully complete the entire imaging session in 2 minutes, which was typical among masked subjects. The increase in time compared to that of the non-masked subject can be attributed to
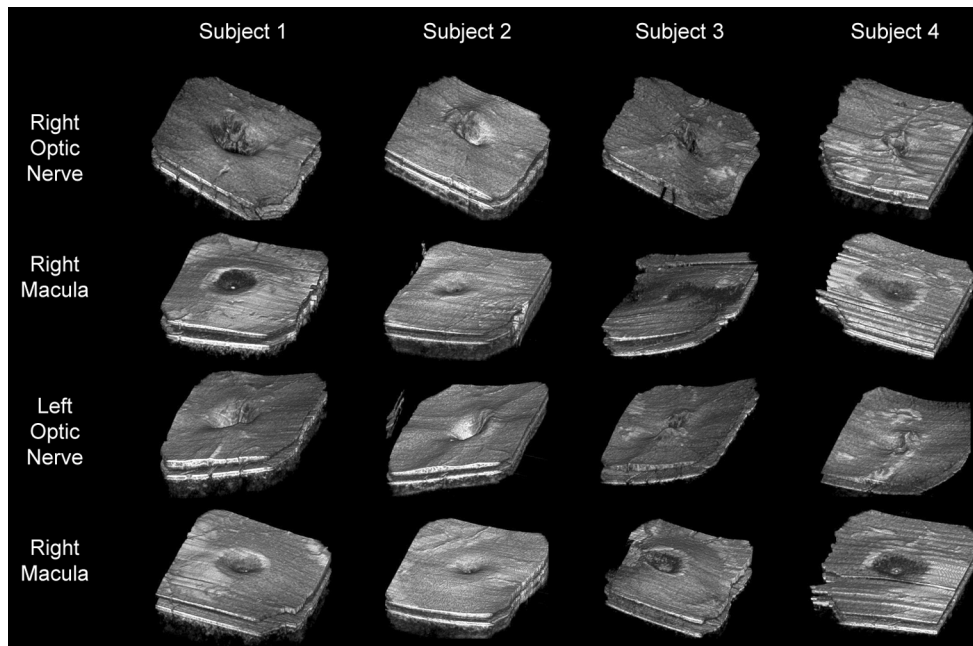
Fig. 9. Volumetric retinal imaging results in free-standing human subjects without a fixation target, mechanical head stabilization, or pupil dilation. The lateral extent of the volumes is $20° \times 20°$.
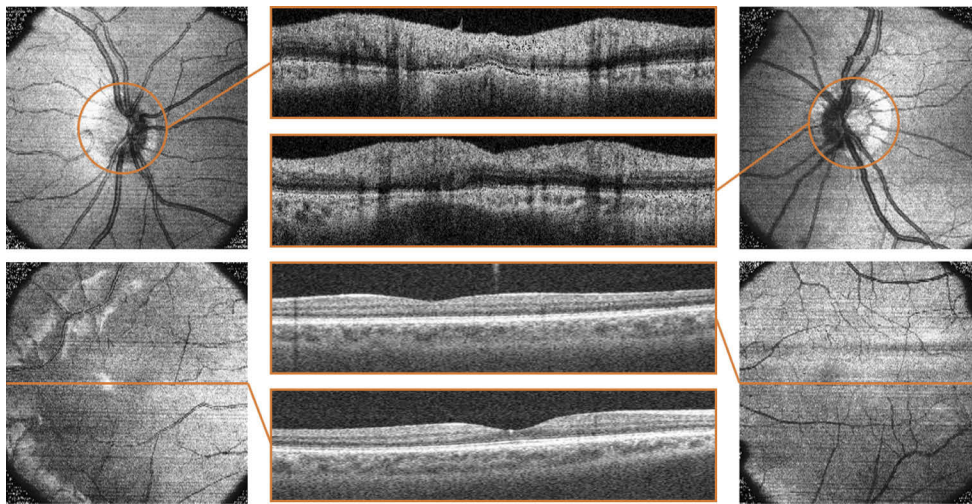

Fig. 10. Retinal scans from a free standing subject. Shown in the top panels are *enface* projections for both left (Upper Left) and right (Upper Right) Optic Nerves, and their corresponding circumpapillary B-Scans. Shown in the bottom panels are *enface* projections for both left (Bottom Left) and right (Bottom Right) maculas, and their corresponding foveal B-Scans. The extent of the *enface* panels is $20° \times 20°$. The lateral and axial extents shown of the B-Scans are $20°$ and $820 \mu m$, respectively.

the time spent by the operator directing the subject to orient their face to optimize the quality of face tracking as well as the time spent adjusting the angle of entrance of the scan to align with each region of interest.

4. Discussion

This manuscript describes the design of a 5 degree of freedom auto-aligning robotical OCT scanner that allowed motion-corrected retinal imaging at multiple regions of interest without the need for a fixation target or mechanical head stabilization. This was achieved using face and pupil cameras for motion tracking, as well as a robotic arm and dynamic optical components for alignment control.

The advantage of the dual-tracking system with face and pupil tracking cameras was that we could image over the extended range of the Face Tracking system while aligning with the accuracy and precision of the pupil tracking system simultaneously. Similarly, the mechanical alignment of the robot and the optical alignment using ray aiming work simultaneously and redundantly to align rapidly and accurately over an extended workspace.

The data presented in this paper demonstrates that this system is capable of imaging multiple regions of interest at the retina, within a 28° angular range, without the need for skilled operator intervention, mechanical head stabilization, or a fixation target. With these capabilities, we envision that a robotically aligned OCT scanner could be used in routine screening roles where specialist ophthalmic photographers are not available. Additionally, such a system and potential extensions could be used to image patients who are not able to utilize a chin rest, such as those who are unconscious, bedbound, or injured, whether they are standing, seated, or supine. Gaze tracking provided the ability to image multiple regions of interest at the retina. This allowed for an automated acquisition of foveal and circumpapillary scans for both eyes, which are the clinically relevant B-scans most often used in OCT eye exams.

The requirement of masking degraded the aligning ability of the system, as it negatively impacted the quality of face tracking. This impact led to the need for semi-autonomous imaging instead of fully-autonomous imaging as well as an increase in imaging time of approximately one minute. However, it is encouraging that the system was still able to auto-align with the subjects and image all relevant regions of interest even under these adversarial circumstances.

The ability to automate the acquisition of clinically relevant retinal scans may increase the accessibility of retinal OCT to non-specialist settings. Primary care settings, for example, could see a new role for OCT in routine physical examinations that are otherwise performed with direct ophthalmoscopy, which does not provide the high-quality volumetric data that OCT scans provide. High quality retinal imaging in routine primary care eye exams could lead to early detection of common eye conditions, such as age-related macular degeneration and glaucoma, on patients getting a annual eye exam. Since the RAOCT system could run without operator supervision, it can even be deployed in more public environments, such as malls, where people can walk to the system and get imaged, even without a scheduled appointment or operator intervention.

An additional advantage of a pupil-tracking OCT scanner is the ability to select an arbitrary pupil entrance position from which to image the subject. The pupil entrance position is responsible for the observed tilt in cross sectional images [51]. In Fig. 9, the angular tilt for a given region of interest is consistent across multiple subjects, demonstrating the ability to control for pupil entry position. For this paper, the pupil entry position was designed to be the geometric center of the pupil, to minimize iris clipping. However, future robotically-aligned OCT systems could leverage the ability to image at arbitrary pupil entry positions. Knowledge of the pupil entry position as well as scan location provides information about the angle of incidence at the retinal plane. Because there are retinal structures, such as Henle's fiber layer, that exhibit illumination directionality dependence [51,52], our system may facilitate studies that exploit this dependency by automating the translation and rotation of the scan pivot and mitigating motion artifacts. Additionally, control over the angle of incidence at the pupil and retinal plane provides angular diversity that may be used for image fusion [53], as well as mosaicking [54].

Disclosures. PO, MD, RPM, ANK, JAI: Duke University (P), JAI: Leica Microsystems (P,R)

Data availability. Data underlying the results in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," Science **254**(5035), 1178–1181 (1991).
2. J. A. Izatt, M. R. Hee, E. A. Swanson, C. P. Lin, D. Huang, J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto, "Micrometer-scale resolution imaging of the anterior eye in vivo with optical coherence tomography," Arch. Ophthalmol. **112**(12), 1584–1589 (1994).
3. S. Radhakrishnan, A. M. Rollins, J. E. Roth, S. Yazdanfar, V. Westphal, D. S. Bardenstein, and J. A. Izatt, "Real-time optical coherence tomography of the anterior segment at 1310 nm," Arch. Ophthalmol. **119**(8), 1179–1185 (2001).
4. E. A. Swanson, J. A. Izatt, M. R. Hee, D. Huang, C. Lin, J. Schuman, C. Puliafito, and J. G. Fujimoto, "In vivo retinal imaging by optical coherence tomography," Opt. Lett. **18**(21), 1864–1866 (1993).
5. J. P. Ehlers, Y. S. Modi, P. E. Pecen, J. Goshe, W. J. Dupps, A. Rachitskaya, S. Sharma, A. Yuan, R. Singh, P. K. Kaiser, J. L. Reese, C. Calabrise, A. Watts, and S. K. Srivastava, "The discover study 3-year results: feasibility and usefulness of microscope-integrated intraoperative oct during ophthalmic surgery," Ophthalmology **125**(7), 1014–1027 (2018).
6. U. Schmidt-Erfurth, S. Klimscha, S. Waldstein, and H. Bogunović, "A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration," Eye **31**(1), 26–44 (2017).
7. J. S. Schuman, C. A. Puliafito, J. G. Fujimoto, and J. S. Duker, *Optical Coherence Tomography of Ocular Diseases* (Slack, 2004).
8. W. Jung, J. Kim, M. Jeon, E. J. Chaney, C. N. Stewart, and S. A. Boppart, "Handheld optical coherence tomography scanner for primary care diagnostics," IEEE Trans. Biomed. Eng. **58**(3), 741–744 (2011).
9. C. D. Lu, M. F. Kraus, B. Potsaid, J. J. Liu, W. Choi, V. Jayaraman, A. E. Cable, J. Hornegger, J. S. Duker, and J. G. Fujimoto, "Handheld ultrahigh speed swept source optical coherence tomography instrument using a mems scanning mirror," Biomed. Opt. Express **5**(1), 293–311 (2014).
10. D. Nankivil, G. Waterman, F. LaRocca, B. Keller, A. N. Kuo, and J. A. Izatt, "Handheld, rapidly switchable, anterior/posterior segment swept source optical coherence tomography probe," Biomed. Opt. Express **6**(11), 4516–4528 (2015).
11. F. LaRocca, D. Nankivil, T. DuBose, C. A. Toth, S. Farsiu, and J. A. Izatt, "In vivo cellular-resolution retinal imaging in infants and children using an ultracompact handheld probe," Nat. Photonics **10**(9), 580–584 (2016).
12. A.-H. Dhalla, R. P. McNabb, P. Ortiz, M. Jackson-Atogi, G. Waterman, J. A. Izatt, and A. N. Kuo, "Hand-held high-speed whole-eye oct: Simultaneous ssoct of the anterior segment and retina using a compact probe," Investigative Ophthalmology & Visual Science **60**, 1295 (2019).
13. C. Viehland, X. Chen, D. Tran-Viet, M. Jackson-Atogi, P. Ortiz, G. Waterman, L. Vajzovic, C. A. Toth, and J. A. Izatt, "Ergonomic handheld oct angiography probe optimized for pediatric and supine imaging," Biomed. Opt. Express **10**(5), 2623–2638 (2019).
14. S. Song, K. Zhou, J. J. Xu, Q. Zhang, S. Lyu, and R. Wang, "Development of a clinical prototype of a miniature hand-held optical coherence tomography probe for prematurity and pediatric ophthalmic imaging," Biomed. Opt. Express **10**(5), 2383–2398 (2019).
15. J. D. Malone, M. T. El-Haddad, S. S. Yerramreddy, I. Oguz, and Y. K. K. Tao, "Handheld spectrally encoded coherence tomography and reflectometry for motion-corrected ophthalmic optical coherence tomography and optical coherence tomography angiography," Neurophotonics **6**(04), 1–11 (2019).
16. H. C. Hendargo, R. Estrada, S. J. Chiu, C. Tomasi, S. Farsiu, and J. A. Izatt, "Automated non-rigid registration and mosaicing for robust imaging of distinct retinal capillary beds using speckle variance optical coherence tomography," Biomed. Opt. Express **4**(6), 803–821 (2013).
17. S. Ricco, M. Chen, H. Ishikawa, G. Wollstein, and J. Schuman, "Correcting motion artifacts in retinal spectral domain optical coherence tomography via image registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer, 2009), pp. 100–107.
18. M. F. Kraus, B. Potsaid, M. A. Mayer, R. Bock, B. Baumann, J. J. Liu, J. Hornegger, and J. G. Fujimoto, "Motion correction in optical coherence tomography volumes on a per a-scan basis using orthogonal scan patterns," Biomed. Opt. Express **3**(6), 1182–1199 (2012).
19. G. Gelikonov, P. Shilyagin, S. Y. Ksenofontov, D. Terpelov, V. Gelikonov, and A. Moiseev, "Numerical method for axial motion correction in optical coherence tomography," in *Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XXIV*, vol. 11228 (International Society for Optics and Photonics, 2020).
20. Z. Chen, Y. Shen, W. Bao, P. Li, X. Wang, and Z. Ding, "Motion correction using overlapped data correlation based on a spatial-spectral encoded parallel optical coherence tomography," Opt. Express **25**(6), 7069–7083 (2017).
21. A. Montuoro, J. Wu, S. Waldstein, B. Gerendas, G. Langs, C. Simader, and U. Schmidt-Erfurth, "Motion artefact correction in retinal optical coherence tomography using local symmetry," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2014), pp. 130–137.
22. N. D. Shemonski, F. A. South, Y.-Z. Liu, S. G. Adie, P. S. Carney, and S. A. Boppart, "Computational high-resolution optical imaging of the living human retina," Nat. Photonics **9**(7), 440–443 (2015).

23. P. J. Rosenfeld, M. K. Durbin, L. Roisman, F. Zheng, A. Miller, G. Robbins, K. B. Schaal, and G. Gregori, "Zeiss angioplex™ spectral domain optical coherence tomography angiography: technical aspects," Dev. Ophthalmol. **56**, 18–29 (2016).

24. S. Orlowski, T. Dziubak, J. Szatkowski, P. Dalasinski, and M. Pańkowiec, "Retinal movement tracking in optical coherence tomography," US Patent 9, 750, 403 (2017).

25. T. Fujimura, T. Kogawa, R. Morishima, H. Okada, and T. Hayashi, "Ophthalmologic apparatus," US Patent 9, 526, 416 (2016).

26. M. Draelos, P. Ortiz, R. Qian, C. Viehland, R. McNabb, K. Hauser, A. N. Kuo, and J. A. Izatt, "Contactless optical coherence tomography of the eyes of freestanding individuals with a robotic scanner," >Nature Biomedical Engineering **5**, 726–736 (2021).

27. M. Draelos, P. Ortiz, R. Qian, B. Keller, K. Hauser, A. Kuo, and J. Izatt, "Automatic optical coherence tomography imaging of stationary and moving eyes with a robotically-aligned scanner," 2019 International Conference on Robotics and Automation **2019**, 8897–8903 (2019).

28. Z. R. Cherif, A. Nait-Ali, J. Motsch, and M. Krebs, "An adaptive calibration of an infrared light device used for gaze tracking," in *IMTC/2002, Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 00CH37276)*, vol. 2 (IEEE, 2002), pp. 1029–1033.

29. X. L. Brolly and J. B. Mulligan, "Implicit calibration of a remote gaze tracker," in 2004 Conference on Computer Vision and Pattern Recognition Workshop (IEEE, 2004), pp. 134.

30. C. Jian-nan, Z. Chuang, Y. Yan-tao, L. Yang, and Z. Han, "Eye gaze calculation based on nonlinear polynomial and generalized regression neural network," in *2009 Fifth International Conference on Natural Computation*, vol. 3 (IEEE, 2009), pp. 617–623.

31. P. Blignaut, "Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker," J. Eye Mov. Res. **7**(1), 1 (2013).

32. D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," Comput. Vis. Image Underst. **98**(1), 25–51 (2005).

33. F. L. Coutinho and C. H. Morimoto, "Augmenting the robustness of cross-ratio gaze tracking methods to head movement," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, (2012), pp. 59–66.

34. A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," IEEE Access **5**, 16495–16519 (2017).

35. M. Draelos, Q. Qiu, A. Bronstein, and G. Sapiro, "Intel realsense= real low cost gaze," in *2015 IEEE International Conference on Image Processing (ICIP)*, (IEEE, 2015), pp. 2520–2524.

36. I. Bacivarov, M. Ionita, and P. Corcoran, "Statistical models of appearance for eye tracking and eye-blink detection and measurement," IEEE Trans. Consumer Electron. **54**(3), 1312–1320 (2008).

37. L. Cao, C. Gou, K. Wang, G. Xiong, and F.-Y. Wang, "Gaze-aided eye detection via appearance learning," in *2018 24th International Conference on Pattern Recognition (ICPR)*, (IEEE, 2018), pp. 1965–1970.

38. I. F. Ince and J. W. Kim, "A 2d eye gaze estimation system with low-resolution webcam images," EURASIP J. Adv. Signal Process. **2011**(1), 40 (2011).

39. W. Wang, Y. Huang, and R. Zhang, "Driver gaze tracker using deformable template matching," in *Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety*, (IEEE, 2011), pp. 244–247.

40. T. Nagamatsu, Y. Iwamoto, J. Kamahara, N. Tanaka, and M. Yamamoto, "Gaze estimation method based on an aspherical model of the cornea: surface of revolution about the optical axis of the eye," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, (2010), pp. 255–258.

41. J. Chen, Y. Tong, W. Gray, and Q. Ji, "A robust 3d eye gaze tracking system using noise reduction," in *Proceedings of the 2008 symposium on Eye tracking research & applications*, (2008), pp. 189–196.

42. E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," IEEE Trans. Biomed. Eng. **53**(6), 1124–1133 (2006).

43. T. Kroger, "Online trajectory generation: Straight-line trajectories," IEEE Trans. Robot. **27**(5), 1010–1016 (2011).

44. O. M. Carrasco-Zevallos, D. Nankivil, C. Viehland, B. Keller, and J. A. Izatt, "Pupil tracking for real-time motion corrected anterior segment optical coherence tomography," PLoS One **11**(8), e0162015 (2016).

45. R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration," IEEE Trans. Robot. Automat. **5**(3), 345–358 (1989).

46. S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," Pattern Recognit. **47**(6), 2280–2292 (2014).

47. T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, (2018), pp. 59–66.

48. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Machine Intell. **22**(11), 1330–1334 (2000).

49. G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools (2000).

50. P. Thévenaz, U. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," IEEE Trans. on Image Process. **7**(1), 27–41 (1998).

51. B. J. Lujan, A. Roorda, R. W. Knighton, and J. Carroll, "Revealing henle's fiber layer using spectral domain optical coherence tomography," Invest. Ophthalmol. Vis. Sci. **52**(3), 1486–1492 (2011).

52. W. S. Stiles and B. Crawford, "The luminous efficiency of rays entering the eye pupil at different points," Proc. R. Soc. Lond. B. **112**(778), 428–450 (1933).

53. K. C. Zhou, R. Qian, S. Degan, S. Farsiu, and J. A. Izatt, "Optical coherence refraction tomography," Nat. Photonics **13**(11), 794–802 (2019).

54. F. Schwarzhans, S. Desissaire, S. Steiner, M. Pircher, C. K. Hitzenberger, H. Resch, C. Vass, and G. Fischer, "Generating large field of view en-face projection images from intra-acquisition motion compensated volumetric optical coherence tomography data," Biomed. Opt. Express **11**(12), 6881–6904 (2020).